



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2019

**Binäre Codes und komplexes Denken: digital Humanities und
Geschichtswissenschaft**

Tanner, Jakob

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-178483>

Book Section

Published Version

Originally published at:

Tanner, Jakob (2019). Binäre Codes und komplexes Denken: digital Humanities und Geschichtswissenschaft. In: Schröter, Juliane. Linguistische Kulturanalyse. Berlin, Boston: De Gruyter, 91-110.

Jakob Tanner

Binäre Codes und komplexes Denken

Digital Humanities und Geschichtswissenschaft

Abstract: Die *Great digitization* hat in den Kultur- und Geisteswissenschaften eine intensive Diskussion um die *Digital Humanities* ausgelöst. Wichtige Inputs kommen dabei von der Korpuslinguistik, die schon früh vom *computational turn* erfasst wurde; mit den hier entwickelten Ansätzen lassen sich die reduktionistischen Ambitionen der *Culturomics* nicht nur kritisieren, sondern auch kreativ nutzen. Der Aufsatz evaluiert die in den 1960er Jahren aufkommende sog. „Begriffsgeschichte“ und bezieht diese auf neue Forschungstools wie Google Ngram-Viewer. Es wird gefragt, wie sich eine sterile Entgegensetzung von alingistischen Analyseverfahren und traditionellen hermeneutischen Methoden vermeiden lässt. Über forschungspraktische Probleme hinaus steht hier immer auch die Definition von Sprache und Kultur zur Disposition. In einer systematischen Verschränkung von *Big Data* und historischer Anthropologie sieht der Verfasser eine vielversprechende Perspektive.

Keywords: Great Digitization, Digital Humanities, Big Data, historische Semantik, Korpuslinguistik, historische Anthropologie

1 The Great Digitization

Die Digitalisierung ist eine „stille Revolution“ (Bunz 2012). Von ihren Auswirkungen her wird sie mit „groß“ in Verbindung gebracht. So in den Wortverbindungen „Great Digitization“ oder „Big Data“.¹ Auch in der Geschichtswissenschaft stoßen Konzepte, die auf große Veränderungen abzielen, weiterhin auf Resonanz. Zu nennen sind etwa „The Great Transformation“ (der Übergang von traditionellen zu Marktgesellschaften), „The Great Divergence“ (die zunehmende Kluft in der globalen Reichumsverteilung seit dem 18. Jahrhundert), „The Great Compression“ (der Abbau der Einkommens- und Vermögensungleichheit in wirtschaftlich

¹ Aus kritischer Perspektive: Polastron 2009; Reichert 2014.

entwickelten Ländern vom Ersten Weltkrieg bis in die 1970er Jahre) oder „The Great Depression“ (die Weltwirtschaftskrise der 1930er Jahre).

„The Great Digitization“ und „Big Data“ haben allerdings eine andere Qualität als die genannten Großbeschreibungen historischer Prozesse. Hier geht es nicht einfach um Forschungshypothesen und -resultate, sondern darum, wie die wissenschaftliche Arbeit organisiert ist und wie sie praktisch funktioniert. Gerade auf dieser pragmatischen Ebene lässt sich ein nachhaltiger Wandel beobachten, weit über die Geschichtswissenschaft hinaus. Die Invasion digitaler Techniken in alle Wissenschaftsdisziplinen hat bisherige Selbstverständnisse von Fächern und eingespielte Forschungsstile im Mark erschüttert. In den Naturwissenschaften breiten sich neue Strategien einer *Data-Driven Research* aus. Die Geisteswissenschaften erfinden sich neu als *Digital Humanities* (Jannidis et al. 2017). Diese Begriffsschöpfung, in der Wissenschaftstraditionalisten ein Oxymoron sehen, betrifft Theorieansätze, Erklärungsmodelle und Erzähltechniken ebenso wie die Skaleneffekte, welche ein sprunghafter Zuwachs leicht verfügbarer „Daten“ unvermeidlich auslöst. In diesen Datenmassiven entfaltet der *computational turn* (Berry 2011) seit einigen Jahrzehnten sein Drehmoment und inzwischen hat sich die Digitalisierung von Forschungsressourcen und Archivdokumenten auf breiter Front durchgesetzt, dies verbunden mit der Anwendung von Grossrechnern für die Auswertung rasant wachsender maschinenlesbarer Quellenkorpora.

Diese Entwicklung reicht in die frühe Nachkriegszeit zurück. Zu den Pionierinnen gehörten zum einen die *Cliometrics*, ein quantifizierender Ansatz in der Geschichtswissenschaft, der seit den 1950er Jahren auf die Erfassung und Auswertung serieller Quellen über längere Zeiträume hinweg abzielte (vgl. Lyons 2008; Diebolt & Hauptert 2016). Zum andern waren es Korpuslinguistik und theologische Komparatistik, die schon früh die Chancen steigender Speicher- und Rechenkapazitäten erkannt und im Computer einen gleichsam natürlichen Verbündeten gesehen haben. So wird in der *Corpus-Linguistics* die B.C.-Ära (*Before Computer*) von jener danach (ab den frühen 1960er Jahren) unterschieden.²

Zu beobachten ist eine Doppelbewegung. In der einen Richtung gelangen immer mehr Computer in die Welt, so auch in den Forschungsbetrieb. Sie legen nicht nur neue Methoden nahe, sondern verändern die Zeitökonomie, d. h. die Opportunitätskosten bestimmter Vorgehensweisen. Ohne Computer und Internet ist wissenschaftliche Forschung längst unmöglich geworden, unabhängig vom Grad der Ablehnung, der Skepsis, der Zustimmung oder der euphorischen Begeisterung, der den digitalen Informations- und Kommunikationstechnologien entgegengebracht wird.

² „Corpus Linguistics has evolved in tandem with computer technology“ (Jensen 2014: 117).

Die Bewegung in die andere Richtung ist weit wichtiger und wird in ihrer Wirkung unterschätzt. Die rasch rechnenden Maschinen können sich nur deshalb in der Gesellschaft ausbreiten, weil die Welt „in den Computer kam“ (Gugerli 2018). Dieser unspektakuläre Vorgang hat weitreichende Auswirkungen. Das Wissen der Menschen gerät unter Veränderungsdruck. Der „Typographic Man“ (McLuhan 1962), der seit dem 16. Jahrhundert die gute alte „Gutenberg-Galaxie“ mit ihren Abermillionen von Büchern bevölkert hatte, ist durch neue Subjektivierungsweisen abgelöst worden. Die Forschung gewöhnt sich daran, dass sperrige und singuläre Objekte – von Urkunden und Drucksachen über unterschiedlichste Bildformate bis hin zu Film- und Tonträgern – die in Archiven, Bibliotheken, Museen und Sammlungen ruhen und dort eingesehen bzw. ausgeliehen werden konnten, zu einem Gutteil durch Digitalisate substituiert werden. Dies kommt einer radikalen *reductio ad binarium* gleich. Das heißt: Alle diese Dokumente werden technisch auf einen Binärcode reduziert. Sie werden verfahrenspraktisch zu homogenen Gütern, die genau gleich gespeichert, übertragen, vervielfältigt oder verändert werden können. Digitalisierung ist allerdings kein Endziel, sondern eine die Effizienz und die Zugänglichkeit bzw. Verfügbarkeit steigernde Umweg-Produktion. Der ganze kulturelle Reichtum einer Gesellschaft wird heruntergebrochen auf Information, die mit nur einer einzigen Unterscheidung – eben binär – operiert. Anschließend ermöglichen diese optimal vereinfachten Binärstoffe dann wiederum oft qualitätsgesteigerte und komplexe analoge Lese-, Seh- und Hörerlebnisse. Im Reich der Töne wird die Verbesserung des „analogen Klangfeelings“ heute (fast) durchwegs digital unterstützt. Und bei digitalisierten Dokumenten (ob Bild oder Text) verschwindet *Le goût de l'archive* (so der Titel einer Studie von Arlette Farge (1989)). Was bleibt, ist ein mono-haptischer Gebrauch von Forschungsressourcen am Bildschirm.

Ein Beispiel stellt die *Venice Time Machine* dar, die aus der Voll-Digitalisierung der über 80 Kilometer Akten, welche die ganze Geschichte Venedigs dokumentieren, hervorgehen soll. Mit einer neuen Generation von Scannern und selbst lernenden Texterkennungsverfahren sollen die im ehemaligen Kloster *Santa Maria Gloriosa dei Frari* (dem Staatsarchiv) von Venedig gelagerten unterschiedlichsten Dokumententypen – von Katasterdaten, Verkehrswegen und Architektur über (Sklaven-)Handel, Vermögenstransaktionen und Alltagsleben bis hin zu politischen Vorgängen, Festivitäten und Kunstwerken – in einen immensen digitalen Datenpool integriert werden, der nicht nur traditionelle Forschungsmethoden erleichtert, sondern es mit spektakulären animierten Visualisierungsmethoden erlaubt, die Stadtentwicklung publikumswirksam im Bild in Szene zu setzen.³

³ Das Projekt unter der Leitung von Prof. Frédéric Kaplan and Dr. Isabella di Lenardo wird von

Technikbegeisterte sprechen bereits von „einer Art Google der Vergangenheit“ (Aebischer 2018). Dieses und weitere Projekte machen deutlich, dass hier eine Geschichte der technischen Traktierung und Neuformatierung kultureller Phänomene im Gange ist, welche im Endeffekt alles, was vorliegt, transformiert, d. h. rechenbar macht. Diese Maschinenpotenz öffnet dann auch dem (forschenden) Menschen neue Möglichkeiten.

Seit die *Digital Humanities* Konturen angenommen haben, gibt es Kontroversen über die Unvermeidlichkeit und Gestaltbarkeit, die Vor- und Nachteile, die Chancen und Gefahren dieses Umbruchs.⁴ Dabei ist nicht einmal klar, ob es sich um etwas grundstürzend Neues handelt oder ob eher ein beschleunigter Wandel des Wissenschaftssystems festzustellen ist, das sich schon vorher immer verändert hat. Ebenso strittig ist, ob der Impetus der „Great Digitization“ die Grenzen der etablierten Fachdisziplinen sprengen wird oder ob hier bloß neue Formen der (alt bekannten) Interdisziplinarität erprobt werden. Weitere Streitpunkte betreffen die Neuausrichtung der Ausbildung, den Umbau von Studien- und Forschungsverfahren und den Ausbau zentraler Rechenkapazitäten mittels Hochleistungscomputern. Ein gewisser Konsens existiert hingegen darüber, dass *Digital Humanities* nicht nur einfach ein neues „Angebot“ bereitstellen, sondern den Wandel, der in den Geistes-, Kultur- und Sozialwissenschaften stattfindet, reflektieren müssen.

2 Google Ngram-Viewer und Culturomics

Die Digitalisierung bietet heute Möglichkeiten, die früher kaum vorstellbar waren. So steht heute für rasche Recherchen der *Google Ngram-Viewer* zur Verfügung.⁵ Auf einem Diagramm zeigt dieser z. B. an, dass der Neologismus „digital humanities“ in der *Anglosphere* erst 1994 auftaucht und dann ein Hype-Verlaufsmuster durchläuft: rascher Anstieg bis 2007, daraufhin ein Einbruch auf ungefähr den halben Wert. Im Deutschen hat sich der Anglizismus im Bereich der Wissenschaftskommunikation zwar etabliert; nichtsdestotrotz bringt der *Ngram-Viewer* keine Kurve; die Bezeichnung ist in diesem Sprachraum statistisch schlicht noch nicht signifikant.

der Eidgenössisch Technischen Hochschule in Lausanne vorangetrieben und versucht, europäische Forschungsgelder zu mobilisieren. Vgl. <https://vtm.epfl.ch/> (23. August 2018).

⁴ Eine besonnen-kritische Position vertritt Hagner (2015).

⁵ Für eine frühe, problemsensible und prospektive Auseinandersetzung mit dem *Google Ngram-Viewer* vgl. Sarasin 2012.

Solche Resultate, die sich für jegliche Worte bis weit zurück vor die Zeit der Aufklärung gewinnen lassen, imponieren und polarisieren. Denn hier wird eine neue Version einer *Armchair-Wissenschaft* angeboten. „Sitting in the comfort of one's home“ (Bell 2018: 75) ist es nun möglich, in Sekundenschnelle eine algorithmengesteuerte Volltextsuche in „lots of books“⁶ durchzuführen, die mit *Optical Character Recognition* (OCR) erschlossen worden sind.

Doch was sagen solche Resultate aus? Wie lässt sich verhindern, dass sich eine falsche Evidenz einstellt? Zunächst durch die Einsicht, dass Algorithmen nicht mit Sprache oder mit Worten, sondern alingustisch arbeiten (vgl. Bubenhöfer & Dreesen 2018: 63). Der *Google Ngram-Viewer* „liest“, wie der Name sagt, N-Gramme ein. Das sind in Fragmente zerlegte Texte, wobei die Grundeinheit ein Zeichen (Buchstabe, Zahl, Leerschlag, etc.) ist. *N* steht für eine Zahl. Es gibt (aufsteigend) Mono-, Di-, Tri-, Tetra-, Penta-, Hexa- und dann Multigramme beliebiger Zahl. Solche Buchstabenfolgen, die – wie in „Digital Humanities“ – als integraler Teil des N-Gramms auch *Blanks* enthalten können, werden in einer Volltextsuche identifiziert.⁷ Es lassen sich für jeden Suchvorgang beliebige, durch Komma voneinander getrennte, N-Gramme eintippen.

Die Präsentation ist diagrammatisch (vgl. dazu Krämer 2016). Das Kurvendia-gramm zeigt horizontal die Zeit und vertikal die relative Auftretenshäufigkeit des entsprechenden N-Gramms an. Dargestellt werden gleitende Jahresmittel, die aus der Zeitreihenanalyse bekannt sind und mit denen höhere Frequenzanteile entfernt werden können; der *Smoothing*-Faktor gibt an, wie viele Werte vor und nach dem Stichjahr in die Durchschnittsbildung einbezogen werden.⁸ Der Suchalgorithmus arbeitet sich durch das „Google Books“-Korpus. Welche Bestände ein- und welch ausgeschlossen sind, wird nicht mitgeteilt. Das Herzstück des Systems ist also eine veritable Black Box, die insofern etwas gelüftet wird, als auf der horizontalen Zeitachse für bestimmte Phasen einschlägige Bücher aufgelistet werden. Diese sind dann wiederum zumindest teilweise als voll durchsuchbare Digitalisate zugänglich.

2009 als neues Recherchetool lanciert, wurde der *Google Ngram-Viewer* 2010 in der Zeitschrift *Science* vorgestellt und 2011 für die Forschung freigeschaltet.⁹

⁶ Der Search-Button lautet schlicht: „Search lots of books“, vgl. <https://books.google.com/ngrams> (6. Februar 2019).

⁷ In der Korpus- und Computerlinguistik spricht man von verschiedenen Verfahren der Tokenisierung. Eine Diskussion der N-Gramm-Bildung für den deutschen Sprachraum findet sich in Schönplüg 1969.

⁸ Beim (voreingestellten, jedoch veränderbaren) Faktor drei sind das insgesamt sieben Werte.

⁹ Vgl. Michel et al. 2011. Das Projekt basierte auf einem „corpus of digitized texts containing about 4 % of all books ever printed“ (Michel et al. 2011: 176).

In den vergangenen Jahren wurde das Tool laufend weiterentwickelt.¹⁰ Die 2004 gegründeten *Google Books* haben gewaltig expandiert. Die Zahl der gescannten Bücher übertrifft alles, was je zuvor für Forschung zur Verfügung stand; waren es 2012 ca. 5 Millionen und 2015 ca. 10 Millionen, so sind es heute bereits ca. 20 Millionen Bücher, was etwa 15 Prozent der um die 130 Millionen weltweit verfassten Bücher entspricht.¹¹ Entscheidende Weichenstellungen in der Konstruktion der Korpora wurden 2012 vorgenommen. Inzwischen lassen sich vier englische Korpora absuchen, dazu Chinesisch, Französisch, Deutsch, Italienisch, Russisch, Spanisch und Hebräisch. Die Sprachenauswahl weist einen ausgeprägten eurozentrischen Bias auf. Die verfügbaren Werke umfassen mehrere Jahrhunderte; suchen lässt sich zwischen 1500 und 2008, wobei signifikante Resultate für den Zeitraum zwischen 1800 und 2000 zu erwarten sind. Auch wenn es mittlerweile recht gut funktionierende Fraktur-OCR-Programme gibt, ist die Texterkennung für die Zeit vor 1800 noch immer schwierig; nach 2000 nimmt der Anteil der hoch volatilen *Digital Born Sources* zu, die nicht Teil der Korpora sind.

Das Hauptproblem ist das Arkanum des Unternehmens. Der Algorithmus ist zwar einfach, doch das Korpus bleibt intransparent. Google gibt weder über Umfang noch über Zusammensetzung Auskunft. Die Auflagenzahl von Publikationen, die zentral sein kann für den Widerhall, den politische Begriffe finden, ist kein Thema. Die Datenbasis stellt insbesondere rechtlich eine Grauzone dar. Dies trifft weit über Google hinaus zu. Der rasch expandierende Cyberspace ist insgesamt gespalten. Ein Teil der Riesenmengen an Daten ist frei verfügbar, während ein anderer Teil durch *Copy Rights* geschützt und – als Folge davon – durch *Paywalls* und Passwörter gegen unbefugten Gratiszugriff abgeschottet wird.

Das Ganze lässt sich am besten mit dem Sputnik vergleichen: Noch bevor Überflug- und Weltraumrechte geklärt waren, piepste 1957 der erste von der UdSSR gestartete Satellit aus dem Orbit und schuf damit neue Verhältnisse, die erst später verrechtlicht wurden (ein Prozess, der in vollem Gang ist). Soweit es

¹⁰ So gibt es seit einigen Jahren die Funktion „case-insensitive“, die es ermöglicht zu wählen, ob Groß- und Kleinschreibung berücksichtigt werden soll. Inzwischen ist auch eine „Wildcard“- und eine „Inflection“-Search möglich; es lassen sich bei gleichlautenden Buchstabenfolgen Verben von Substantiven unterscheiden und es gibt verschiedenste Kombinationsvarianten. <https://books.google.com/ngrams/info>; vgl. auch Rosenberg 2017.

¹¹ Ein Buch ist ein Objekt, auf das sich die Metadaten beziehen. Die Definition ist unabhängig von der Auflage; ob Millionen Kopien produziert wurden oder nur eine Handvoll Exemplare bestehen, spielt keine Rolle. Hingegen werden z. B. Hardcover- und Paperback-Ausgaben speziell erfasst. Hingegen zählt ein Buch, das eine Reihe von Pamphleten umfasst, nur als eines. <http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html> (23. August 2018).

heute bei *Google Books* um das *Library Project* geht, so werden ohne Berücksichtigung der Rechtslage laufend *faits accomplis* geschaffen. Nach dem Prinzip des geringsten Widerstandes in Kombination mit ziemlich rohen Repräsentativitätskriterien wird im Massenstil darauf losgescannt. Es gibt allerdings ein *Partner Program*, bei dem die Urheberrechte und die Abmachungen mit den Verlagen geklärt sind.¹² Insgesamt nutzt der *Google Ngram-Viewer* einen doppelten Spielraum: Einerseits die Rechtsunsicherheit betreffend die Verwendung kurzer Exzerpte (sog. „Snippets“ oder „Schnipsel“) und andererseits das Argument, dass die ganze Maschinerie der Forschung dient.¹³

Wie auch immer sich die Rechtslage entwickelt: Das Riesenkorpus von *Google Books* wird in Betrieb bleiben und an inhaltlicher Breite und zeitlicher Tiefe gewinnen. Dass der Bestand, soweit das rechtlich möglich ist, gratis benutzt werden kann, verweist auf die komplexere Kommerzialisierungsstrategie des Unternehmens, das nicht seine Scan-Leistung verkauft, sondern in allen Geschäftsfeldern auf die rentable Verwertung der von NutzerInnen gewonnenen Daten setzt.

Umso wichtiger ist es, dass es Konkurrenz gibt. Eine große Zahl weiterer Plattformen und Bereitsteller digitalisierter Drucksachen, Bücher und Medien aller Art ist teils privatwirtschaftlich organisiert, teils öffentlich oder durch Mäzenaten finanziert. Zu erwähnen ist etwa das 1996 als Web-Archiv gegründete und ab 1999 Bücher und weitere Dokumente, aber auch Software zur Verfügung stellende *Internet-Archive* (San Francisco). Als *Million Book Project* stellt dieses Archiv über eine *Open Library* mittlerweile über 10 Millionen Texte zur freien Verfügung; im Unterschied zu *Google Books* handelt es sich ausschließlich um Publikationen, die gemeinfrei (geworden) sind.

Daneben und darüber hinaus gibt es Dutzende von weiteren Anbietern, teils fokussiert auf Sprachgebiete, teils mit Schwerpunkt auf bestimmte Phasen der Geschichte.¹⁴ Wichtig für die Wissenschaft ist zudem das *Web of Science* mit über 50 Millionen Datenbankeinträgen; in den 1960er Jahren entstanden, wurde diese Forschungsplattform 1992 von der *Thomson-Reuters*-Gruppe übernommen und 2017 an *Clarivate Analytics* weiterverkauft. Erwähnt werden muss auch das der *American Association for the Advancement of Science* (AAAS) gehörende *Science*

¹² <https://www.google.com/intl/en/googlebooks/about/index.html> (23. August 2018).

¹³ Zu den Versuchen, die Verwendung von *Snippets* einzuschränken oder kostenpflichtig zu machen, vgl. Potthast et al. 2018.

¹⁴ Als Beispiel vgl. die *Perseus Digital Library* mit einem Schwerpunkt in der Antike (<http://www.perseus.tufts.edu/hopper/>, 23. August 2018) oder die *Gallica*, die e-Bibliothek der Französischen Nationalbibliothek (http://www.bnf.fr/en/collections_and_services/digital_libraries_gallica.html, 23. August 2018).

Magazine, das Journal-Artikel zugänglich macht.¹⁵ Seit einiger Zeit schon sind „digitale Werkzeuge [auch] in der Mediävistik angekommen“ (Rohmann 2015) und haben hier den Trend verstärkt, Texttechnologie, Editionsphilologie und historische Forschung synergetisch zu betrachten. Die *Computational Historical Semantics*, das *Historical Semantics Corpus Management* und das *Frankfurt Latin Lexicon* (mit 11 Millionen Wortformen) bieten je unterschiedliche Möglichkeiten für korpusbasierte Abfragen in mittellateinischen Quellenbeständen an (vgl. Rohmann 2015).

Google Books versucht, alle diese Projekte mit einer Skalenökonomie zu übertrumpfen. Das Unternehmen verspricht, die Restriktionen, denen kleinere, auf bestimmte Phasen und Forschungsinteressen fokussierte Projekte unterliegen, zumindest für die Neuzeit aufzuheben und das Textangebot in neue Größenordnungen zu steigern. Damit soll einem völlig neuen Nutzungs- und Forschungsstil zum Durchbruch verholfen werden. Im *Science*-Beitrag von 2010, in dem das neue Verfahren der „Quantitative Analysis of Culture Using Millions of Digitized Books“ vorgestellt wurde, war von einer *Culturomics* die Rede. Diese „extends the boundaries of rigorous quantitative inquiry to a wide array of new phenomena spanning the social sciences and the humanities“ (Michel et al. 2011: 176).

Das ist ein hoher Anspruch, der für die Geistes- und Kulturwissenschaften eine Herausforderung darstellt. Es gibt die Meinung, *Culturomics* sei nur der Auftakt zu einem weit fundamentaleren Unternehmen, mit dem sich die Forschung unter Einbezug von *News Media* in die Echtzeit-Dimension hinein entwickeln könne, was neue Verfahren der Vermessung und Erklärung menschlicher Gesellschaften auf globaler Stufenleiter ermöglichen würde. Eine solche Vision entwickelt das Projekt einer *Culturomics 2.0*, welches die retrodigitalisierten Archive hinter sich zu lassen und einen „cross-national real-time record of human society“ zu erschließen verspricht – mit hybriden Erwartungen und der impliziten Botschaft, dass die Sozial- und Geisteswissenschaften, wie wir sie kannten, nur noch Auslaufmodelle sein können (vgl. Leetaru 2011).

3 Geschichtliche Grundbegriffe

In einer solchen Situation sind nur allzu oft polarisierte Reaktionen zu erkennen, die sich häufig am *Google Ngram-Viewer* kristallisieren. Während die einen von

¹⁵ Auch (fast) jedes Journal bietet mittlerweile die veröffentlichten Artikel an; allerdings mit sehr unterschiedlichen „Moving Walls“ und Konditionen.

aufregenden neuen Forschungsperspektiven sprechen, winken andere entnervt ab. Die einen feiern die Fortschritte der Automatisierung. Sie stellen nicht nur das Ende der mühsamen Maloche in widerständigen Analogmedien und der Kärnerarbeit im staubigen Archiv in Aussicht, sondern schwärmen von datengetriebenen Zugängen, welche der Forschung neue, im Rahmen etablierter Theorien kaum erahnte Möglichkeiten erschließen könnten. Die andern hegen Befürchtungen, sei es, dass sie *Big Data* als sinnlose Stoffhuberei betrachten, sei es, dass sie befürchten, die Algorithmen würden forschende Menschen künftig arbeitslos machen.

In einer solchen Lage, in der sich die ganze Diskussion um technische Innovation in der Arbeitswelt widerspiegelt, lohnt sich ein Blick zurück auf ein Vorhaben, das die Historizität von Begriffen, die Ausdifferenzierung und den Wandel ihrer Bedeutung, ins Zentrum stellte: Die *Geschichtlichen Grundbegriffe*. Dieses von Analogmedien ausgehende Projekt wurde in den 1960er Jahren aufgegleist und war dem Ansatz einer politischen Sozialgeschichte verpflichtet. Als „alphabetisch sortiertes Handbuch zur semantischen Geburt der Moderne“ (so Geulen 2010: 80) wollte es Leitbegriffe, Schlüssel- und Schlagworte der „geschichtlichen Bewegung“ verständlich machen, und zwar sowohl in ihrer enormen Persistenz wie auch in ihrem oft volatilen Auf-und-wieder-Abtauchen. „Unser Vorhaben [...] interpretiert die Geschichte durch ihre jeweiligen Begriffe so, wie es die Begriffe geschichtlich versteht“, schrieb Reinhart Koselleck als einer der Herausgeber und *Spiritus Rector* in der Einleitung des 1972 erschienenen ersten Bandes des Werks, das beim Abschluss 1997 acht Bände umfasste (Koselleck 1972: XXIII). Das (im Untertitel so genannte) *Historische Lexikon der politischen und sozialen Sprache* unterstellte eine „Konvergenz von Begriff und Geschichte“ und lehnte eine Vereinseitigung sowohl von Sachverhalten wie auch von Interpretationen ab (Koselleck 1972: XXIII).¹⁶ Sprache wird demnach als Faktor und als Indikator der geschichtlichen Bewegung verstanden.

Um ein solches Unternehmen in Printform durchführbar zu machen, mussten ca. 130 Grundbegriffe ausgewählt werden. Maßgebliches Auswahlkriterium ist die Frage nach Umbruchphasen, die mit einem „tiefgreifenden Bedeutungswandel klassischer topoi“ einherläuft (Koselleck 1972: XV). Die Herausgeber identifizierten die Phase ab 1750, mithin die Aufklärung, als „Sattelzeit“, als eine Zeit beschleunigter, verdichteter Veränderung, in der sich eine alte Welt auflöst und der Übergang zur Moderne stattfindet. Ausgangspunkt ist die „neuzeitliche Begrifflichkeit“ der Zeit um 1700, die in der zweiten Hälfte des 18. Jahrhunderts in den Sog moderner Bedeutungsströme gerät und mit „neuen Sinngehalten“ aufgeladen wird. Damit erhalten die analysierten Begriffe eine Janusköpfig-

¹⁶ Dieses Problem wird auch ausgeführt in Koselleck 1991.

keit: rückwärts betrachtet gehören sie einer untergegangenen Welt an und sind demnach fremd, unverständlich; vorwärts gelesen stellen sie die Vorgeschichte der Gegenwart dar und sind höchst präsent (vgl. Koselleck 1972: XV–XVII). Die Rekonstruktion ihrer Genealogie ermöglicht so eine „semantologische Kontrolle für unseren gegenwärtigen Sprachgebrauch“ (Koselleck 1972: XIX). Die Phase der Bedeutungsumprägung wird nun weiter spezifiziert und dabei stechen vier basale Vorgänge – Demokratisierung, Verzeitlichung, Ideologisierbarkeit und Politisierung – ins Auge. In dieser neuen Konstellation werden auch verstärkt polemische Gegenbegriffe geschmiedet – „Fortschritt“ versus „Reaktion“, „Demokratie“ versus „Autokratie“ etc. – welche den politischen Kampf in einem durch die Pole links versus rechts strukturierten gesellschaftlichen Feld befeuern (vgl. Koselleck 1972: XVI–XVIII).

Eine solche Konzeption bedeutet zweierlei: Erstens basiert sie auf voraussetzungsreichen Fragestellungen, die aus differenzierten und differenzierenden Erkenntnisinteressen hervorgehen. In den Fokus geraten sozio-linguistische und politische Aspekte, die auf territoriale Kategorien (z. B. Frankreich während der Revolutionszeit) bezogen werden. Nicht die Auftretensfrequenz von Buchstabenfolgen, sondern die Symptomatik und die Effekte des Sprachgebrauchs stehen im Zentrum der Aufmerksamkeit. Allen diesen Überlegungen liegt die These einer „sprachlichen Konstitution menschlicher Erfahrung“ bzw. verschiedener „sprachlicher Ermöglichungen von Welterfahrung“ zugrunde (Koselleck 1991: 2).

Zweitens stellten die *Geschichtlichen Grundbegriffe* eine strukturelle Überforderung aller Beteiligten dar. Ohne ein gerütteltes Maß an informierter Willkür hätte sich die Aufgabe gar nicht bewältigen lassen. Reinhart Koselleck schränkt das Vorhaben einleitend deutlich ein, indem er festhält, dass nicht das „gesamte politische und soziale Vokabular“ abgedeckt werden soll und dass das Geleistete höchstens eine „Vorarbeit“ für eine „linguistisch abgestützte politische Semantik“ sein könne (Koselleck 1972: XIV). Die an den „Historischen Grundbegriffen“ Beteiligten sind durchwegs etablierte Historiker und ExpertInnen (der Frauenanteil liegt weit unter 10 Prozent). Verarbeitet wurde vor allem „Höhenkamm-Literatur“, was mit dem Anspruch, die wirklichkeitsformatierenden kognitiven Konzepte darzustellen, korrespondierte. Was nicht berücksichtigt wurde und in welchem Verhältnis die zitierten Texte zur Menge des grundsätzlich verfügbaren Materials stehen, wird nirgends ausgewiesen.

Der *Google Ngram-Viewer*, dessen Algorithmus Zeichenfolgen zählt und relativ gewichtet, verlagert dieses zweite Problem auf die Ebene des Korpus. Dieses kann nach allen möglichen Begriffen durchforstet werden. Die Auswahl wird von den Anbietern auf die Nachfrager verlagert. Und gerade weil *Google Books* entlang der Vektoren „Scan-Kapazität“ und „Zugang“ drauflos arbeitet und das Projekt in Größenordnungen hineintreibt, welche die individuelle Hand- und Kopfarbeit

eines jeden einzelnen Autors – so gut er vernetzt sein mag – übersteigt, fließt mehr Material in die Analyse ein. Diese enorme Verbreiterung der Basis wird allerdings erkaufte durch den völligen Verzicht auf jegliche Fragestellung außer jener, nach welcher der Algorithmus durch das gigantische Korpus hindurch prozessiert und die per se völlig sinnlos ist, weil das System nicht unterscheidet, ob nach einem Schlüsselbegriff oder nach „e e“ gesucht wird.

Nur: Einsichten lassen sich eben auch mit einem Suchsystem, das den maximalen Effort im digitalen Speichern von Texten mit der minimalen Auswertungsroutine verbindet, gewinnen. Der *Google Ngram-Viewer* gibt z. B. Auskunft über auf- und absteigende Trends und ihre Kreuzungspunkte. Im deutschen Sprachraum schneiden sich „Pferd“ und „Auto“ punkto relativer Auftretensfrequenz im Jahre 1964.¹⁷ Genau im selben Jahr kreuzen sich „Energie“ und „Information“. „Maschine“ liegt ab 1983 hinter „Computer“ zurück; „Foucault“ überholt „Sozialgeschichte“ anno 1997,¹⁸ während der „Manager“ noch bis 2003 warten muss, bis er den „Führer“ hinter sich gelassen hat. Soweit so trivial. Die Intention, dass sich diese Begriffe grosso modo gegenläufig entwickeln, bestätigt sich. Die Annahme, dass es sich um reziproke Konzepte handelt, ist allerdings hochgradig erklärungsbedürftig. Viele Gegensatzpaare überschneiden sich gar nie, weil das eine Wort (z. B. „Demokratie“ oder „Wahrheit“) durchwegs häufiger verwendet wurde als ein anderes, das sein Gegenteil sein könnte (in diesem Fall „Führerstaat“ bzw. „Autokratie“ oder „Lüge“). Es ist zudem interessant zu sehen, dass nach 1945 immer weniger von „Reichtum“ und „Not“, längerfristig jedoch mehr über „Armut“ geschrieben wurde und dass letztere im Jahre 1982 „Reichtum“ und 1996 „Not“ überholte – mit anhaltend steigendem Abstand.

Das sind willkürlich ausgewählte Beispiele; ein Vorteil des *Google Ngram-Viewers* besteht darin, dass sich eben beliebige Begriffe eingeben bzw. testen lassen. Die Null-Information zur Semantik und Pragmatik der Begriffe lässt sich in einigen Fällen – in Übereinstimmung mit der These der „Historischen Grundbegriffe“ – überspielen, in dem davon ausgegangen wird, dass der semantische *Basso continuo* von „Demokratie“ im ganzen 20. Jahrhundert grundsätzlich positiv aufgeladen war und dass in der Bekämpfung von Armut zunehmend eine politische Aufgabe erkannt wurde. Eine weiterführende Analyse, die auf Bedeutungsverlagerungen oder auf soziale und geographische Spezifika des Wortgebrauchs schließen lassen könnte, kommt nicht umhin, mit traditionellen Methoden der Textinterpretation zu arbeiten.

17 Alle Google-Ngram-Viewer-Ergebnisse in diesem Text wurden letztmals überprüft am 27. Juni 2018.

18 Die Datenbasis von 2012 wies noch die Jahrzahl 1999 aus (vgl. Sarasin 2012: 151).

4 Historische Semantik revisited

Die Historische Semantik wurde von unterschiedlichen Positionen her kritisiert.¹⁹ Insbesondere die angelsächsische Ideengeschichte (*Conceptual/Intellectual History*) grenzte sich gegenüber dieser Methode ab mit dem Argument, Wörter per se hätten keine Geschichte. So konstatierte Quentin Skinner, der führende Vertreter der *Cambridge School of Intellectual History*, er lege im Unterschied zu Reinhart Koselleck „kein besonderes Gewicht auf die Untersuchung von Begriffen“, sondern sein Forschungsinteresse fokussiere auf „bestimmte Vokabulare“ und darauf, „wie diese in bestimmten diskursiven Situationen zu bestimmten diskursiven Zwecken benutzt wurden“. Auf diese Weise könnten „Texte wieder als die Handlungen“ rekonstruiert werden, „die sie einst, in ihrem ursprünglichen diskursiven Kontext, darstellten“ (Skinner 2009: 16).

Die heiße Phase dieser Konfrontation, welche die deutsche Begriffsgeschichte und die angelsächsische *Cambridge School* beinahe „to the point of beginning a kind of academic war“ brachte (wie sich Lucian Hölscher im Rückblick auf eine 1987 stattfindende Konferenz in New York erinnerte, Hölscher 2002: 3), kühlte sich in der Folge ab. Damit differenzierte sich das Interesse an sprachgeschichtlichen Untersuchungen aus. Insbesondere Autoren, die in der Tradition der Begriffsgeschichte argumentieren, betonten weit stärker die Komplementarität verschiedener Zugänge als deren Unvereinbarkeiten. So hob Christian Geulen die „uneingestandene Wahlverwandtschaft zwischen Kosellecks begriffshistorischer, Foucaults diskursanalytischer und Blumenbergs metaphorologischer Form historischer Semantik“ hervor.²⁰ Ernst Müller (2011) sprach von einer „verspäteten Wirkung“ Kosellecks und zeigt, wie der begriffsgeschichtliche Ansatz seit den 1990er Jahren in einem internationalen Netzwerk auf undogmatische, jedoch durchaus theoretisch ambitionierte Weise erneut aufgegriffen und diskutiert wurde, bevor er dann in der 1998 unter anderem von Melvin Richter und Kari Palonen gegründeten HPSCG (*History of Political and Social Concept Group*) internationalisiert und interdisziplinär ausgerichtet wurde (Müller 2011; vgl. auch Müller & Schmieder 2016).

Sowohl Christian Geulen mit seinem *Plädoyer für eine Geschichte der Grundbegriffe im 20. Jahrhundert* wie Lucian Hölscher mit seinem *European Political Lexicon* insistieren auf der Produktivität einer begriffsgeschichtlichen Perspektive. Hölscher, der auf ein ländervergleichendes und zugleich transnationales Begriffsregister hinarbeitet, argumentiert, Begriffe seien „mobiler“ als der

Diskurs, sie dienten als Schlagworte und könnten als linguistische Einheit isoliert werden, was wiederum ein lexikographisches System in alphabetischer Ordnung unterstütze (vgl. Hölscher 2002: 5).

Christian Geulen argumentiert grundsätzlicher, indem er festhält, die „Satelzeit“ der Aufklärung müsse selbst wiederum historisiert werden, was es ermögliche, die Transformation in die Moderne als Transformation der Moderne weiterzudenken. Er vertritt die Hypothese, dass sich das 20. Jahrhundert als eine weitere „Schwellenzeit“ erweisen könnte, „in der die Grundbegriffe moderner Selbstbeschreibung einen nochmaligen Strukturwandel vollzogen, eine Metamorphose ihrer Semantik und eine Modifikation im Verhältnis ihres Erfahrungs- und Erwartungsgehalts“ (Geulen 2010: 85–86). Dabei sind wiederum vier Vorgänge zu beobachten: Verwissenschaftlichung, Popularisierung, Verräumlichung und Verflüssigung. Zwischen diesen ergeben sich vielfältige Wechselwirkungen: Die Ausweitung des populären Gebrauchs von Begriffen führt zu einem semantischen Konturverlust von Begriffen; und mit der Krise der Fortschrittsideologie in der Wissensgesellschaft werden Erwartungshorizonte zu „Erwartungsräumen“, in denen die horizontale (geographische) Proliferation das vertikale (zeitliche) Noch-Nicht ablösen. Diese Veränderungen betreffen nicht nur die fortgesetzte Aufladung alter Wörter mit neuen Sinngehalten oder die Kreation neuer Begriffe, sondern auch die Materialität der Medien. Im Verlaufe des 20. Jahrhunderts werden diese in einer Weise revolutioniert, dass „schon auf der basalen Ebene lexikalischer Begriffsbestimmungen (...) ein ungleich größeres Quellenkorpus erzeugt“ (Geulen 2010: 96) werden kann (vgl. auch Geulen & Bermes i. V.; Geulen i. Dr.).

Für solche und weitere Projekte kann der *Google Ngram-Viewer* unschätzbare Dienste leisten. Er bietet eine Sonde, mit der ergiebige und zeitsparende Vor-Recherchen durchgeführt werden können, mit denen sich das Feld eines solchen Forschungsvorhabens abstecken lässt. Im Folgenden soll das Potenzial dieses Instruments anhand eines alten („Terror“/ „Terrorismus“) und eines neuen Stichwortkomplexes („Schizophrenie“/„schizophren“) spezifischer ausgelotet werden.

Der Eintrag „Terror/Terrorismus“ von Rudolf Walther in Band 6 der *Historischen Grundbegriffe* aus dem Jahre 1990 liefert eine historische Semantik eines Begriffs, die auf umsichtiger Recherche und enormer Belesenheit basiert und die Bedeutungswandel und differenzierungen auf bemerkenswerte Weise nachzeichnet. Fast drei Jahrzehnte später legt der US-amerikanische Historiker Ronald Schechter eine Studie zur „Genealogie des Terrors im Frankreich des 18. Jahrhunderts“ vor, die, basierend auf einer ganzen Reihe von digitalen Korpora (Google Books, Gallica, ARTEL, French Revolution Digital Archives, Gate's Collection etc.), in minutiöser Weise dem Auftreten des Begriffs *Terreur* nachgeht (Schech-

¹⁹ Einen Überblick gibt Steinmetz (2008).

²⁰ Dies in Anlehnung an Ralf Konersmann (1994), vgl. Geulen 2010.

ter 2018). Zwei Beobachtungen sind erwähnenswert: Zum einen gibt Schechter kaum Auskunft über seine Forschungsmethode (und zieht eine wichtige Studie zum Thema, Crawford 2013, nicht bei); zum andern lassen sich aus seinen Befunden, so gut sie empirisch abgestützt sind, auch ziemlich andere Schlussfolgerungen ziehen. Schechter kehrt die Beschreibung von Walther, wonach „Terreur“ im vorrevolutionären Frankreich negativ konnotiert gewesen, während der Jakobinerherrschaft vom September 1793 bis zum Juli 1794 jedoch zur Staatsdoktrin der „Tugend“ aufgestiegen sei, um. Sein Befund lautet: *Terreur* wurde vor der Französischen Revolution durch alle Verwendungsbereiche (Literatur, Religion, Recht, Medizin, etc.) hindurch als positives Konzept verwendet und erst durch seine blutige Exekution während der revolutionären Transformation ereignete sich schlagartig die Umpolung ins Negative, die bis heute fortwirkt.

In einer Buchbesprechung stellte David A. Bell allerdings fest, dass die positiv-schauerliche Bedeutungsfacette ins 19. Jahrhundert hinein fortwirkte, während die Abwendung von Terror als eines politischen Mittels schon vor der Revolution deutlich festzustellen war (vgl. Bell 2018: 75). Damit nähert er die Interpretation wiederum jener von Rudolf Walther an. Das Beispiel verdeutlicht, dass weniger ein bloßes Abrufen von N-Grammen als der Zugriff auf digitalisierte Korpora, die immer leichter durchsuchbar werden, neue Forschungsmöglichkeiten eröffnet. Zudem wird klar, dass diese neuen Möglichkeiten die anhaltenden Interpretationskonflikte nicht aus der Welt schaffen können. An der Schwerarbeit, welche die Verfasser der *Historischen Grundbegriffe* seit den 1970er Jahren leisteten, führt auch heute kein Weg vorbei. Die Vorstellung, dass Automatisierung und Rationalisierung auf einfache Weise den Aufwand reduziere, stimmt gerade für die Forschung nicht. Die „Great Digization“ macht sie vielmehr tendenziell noch anspruchsvoller.

Der Begriff „Schizophrenie“ gehört vollständig ins 20. Jahrhundert. Im deutschen Sprachraum wurde er seit der Prägung des Begriffs durch Eugen Bleuler in den Jahren 1908 (Vortrag) und 1911 (Handbuch) immer häufiger verwendet, mit einem Höhepunkt Ende der 1930er Jahre. Anschließend folgten Auf- und Abbewegungen, bis ab den 1990er Jahren wiederum eine deutliche Zunahme zu verzeichnen war. Im Französischen taucht der Begriff signifikant weniger häufig auf. Der historische Höchstwert war hier erst in den Jahren um 1960 erreicht; der Wiederanstieg fiel bereits auf die 1980er Jahre. Im Englischen erfolgte der Anstieg ab den 1930er Jahren. In den beginnenden 1980er Jahren wurde ein ähnlich hoher Wert wie im Deutschen fast ein halbes Jahrhundert zuvor erreicht; anfangs des 21. Jahrhunderts ist ein neuer Peak zu beobachten.

„Geisteskrankheit“ ist ein Begriff, der im Deutschen in den 1850er Jahren und dann wiederum – auf weit höherem Niveau – anfangs des 20. Jahrhunderts hohe Werte erreichte, um sich ab 1902 rückläufig zu entwickeln und im Jahre 1920 von

„Schizophrenie“ überholt zu werden. Andere Vergleichsbegriffe wie „Psychopharmaka“ tauchen erst in der Nachkriegszeit auf und erreichen nie die Auftretenshäufigkeit von „Schizophrenie“; umgekehrt ist „Depression“, ein Begriff, der immerzu häufiger auftritt, sich allerdings auch auf wirtschaftliche und andere Entwicklungen bezieht, die sich gegenüber der psychiatrischen Diagnose nicht abgrenzen lassen.

Werden solche Ergebnisse evaluiert, sind zwei Punkte zu beachten. Erstens ist „Schizophrenie“ eine psychiatrische Kategorie, die häufig in wissenschaftlichen Texten vorkommt, die durch *Copyrights* geschützt sind. Google Books hält sich grundsätzlich nicht an solche Regeln; es ist aber zu vermuten, dass diese doch einen Einfluss auf die Struktur der gescannten Texte haben. Zudem reduziert der Aufstieg des Englischen als Wissenschaftssprache seit einigen Jahrzehnten auch in der deutschsprachigen Psychiatrie die Auftretenshäufigkeit des deutschen Terminus.²¹

Zweitens handelt es sich bei den oben kommentierten Kurven um „Forschungsergebnisse“, die ohne eine Fragestellung erzielt wurden, so dass sie inhaltsleer sind. Wer sich für die Sozialtopographie und den Kontext der Wortverwendung interessiert, kommt auf diese Weise nicht weiter. Befunde, wie sie von einer Zürcher Forscherinnengruppe erzielt wurden, wonach für die zweite Hälfte des 20. Jahrhunderts eine Ausweitung der Referenzkontexte und der außerpsychiatrischen Begriffsverwendung festzustellen ist, lassen sich nur durch aufwändigere Forschungsmethoden erzielen, die ebenfalls computergestützt sind und auf digitalisierten Korpora basieren. Bevor sie für Recherchen und Auswertungen zur Verfügung stehen, müssen die Texte gesammelt, aufbereitet, annotiert und für die Analyseprogramme lesbar gemacht werden, was große Vorleistungen nötig macht, wie dies das Dissertationsprojekt von Yvonne Ilg zum Gebrauch von „Schizophrenie“ in der Alltagssprache im Jahrhundert zwischen 1908 und 2009 deutlich macht (Ilg i. V.; vgl. Ilg 2016; Ilg & Maatz 2015; Ilg & Schnedermann 2014). Die Tatsache, dass eine im politischen Zusammenhang verwendete Formulierung wie „schizophrene Situation“ im Google-Diagramm 1951 aus dem Nichts auftaucht und bis 1992 immer häufiger auftritt, um dann bis 2008 auf die Hälfte des Spitzenwerts zurückzufallen, kann so besser verstanden werden.

Eine solche differenzierende Evaluation von Forschungsbereichen kann aus einer schroffen Entgegensetzung traditioneller hermeneutischer Methoden auf

²¹ Tatsächlich zeigen die Google-Ngram-Kurven im Englischen keine Einbrüche wie im Deutschen (schon ab den ausgehenden 1930er Jahren und dann wiederum ab Mitte der 1950er Jahre) und im Französischen (ab anfangs der 1960er Jahre).

der einen Seite und „Culturomics“ auf der anderen herausführen.²² Mittlerweile wurde eine ganze Reihe von Vorschlägen gemacht, wie das ganze „Digital Data Universe“ inklusive der *Social Media* als *Digital Born Sources* für geistes- und kulturwissenschaftliche Projekte genutzt werden kann (vgl. Manovich 2017). Und es gibt Anleitungen, wie sich die Prädominanz des Google-Unternehmens durch den Zugriff auf alternative Datensätze (bibliographische Datensammlungen, Journal-databases etc.) vermindern lässt (vgl. Chumtong & Kaldewey 2017).

Die große Herausforderung der „Great Digitization“ besteht insgesamt weniger in diesem forschungspraktischen Umgang mit riesigen Datenmengen, wie er heute in Korpuslinguistik, historischer Semantik, Diskursanalyse und *Culturomics* gleichermaßen revolutioniert wird. Selbstverständlich ist hier ein professionelles und angesichts eines rasch expandierenden Möglichkeitsraums kreatives Arbeiten gefordert, das immer die prekären Tendenzen zur Privatisierung und kommerziellen Verwertung von Wissen im Auge haben muss. Weit grundsätzlicher geht es allerdings um ein neues Verständnis dessen, was Kultur ist und wie Sprache funktioniert. Der in den 1960er Jahren einsetzende „linguistic war“ zwischen generativer Grammatik und Korpuslinguistik (vgl. Harris 1993) wirft ebenso spannende wie grundsätzliche Fragen auf wie der „Linguistic Turn“ und die „Tropologie des historischen Diskurses“ der 1970er Jahre.²³

Diese Grundsatzfragen hängen mit neuen Erkenntnisinteressen und konzeptionellen Verschiebungen und zugleich mit dem Aufkommen des Computers zusammen. Dieser brach einer von der Vorgehensweise her präzedenzlosen quantitativen Analyse des Sprachgebrauchs Bahn und setzte damit linguistische Theorien und Methoden einem theoretisch-methodischen Stresstest aus. Der Linguist Noah Bubendorfer hat darauf hingewiesen, dass die maschinelle Textverarbeitungsforschung – wie schon am Beispiel des *Google Ngram-Viewer* gezeigt – zunehmend alingustisch verfährt (vgl. Bubenhofer & Dreesen 2018: 63). Würde sich diese Tendenz verallgemeinern, so wäre damit ein prekärer Bedeutungsverlust geisteswissenschaftlicher Forschungstraditionen verbunden (vgl. Bubenhofer & Dreesen 2018: 65). Denn die „statistische Wende in der Computerlinguistik“ (Bubenhofer 2018: 20), d. h. die algorithmische Auswertung natürlicher Sprachen auf formale Muster, auf Regelmässigkeiten und Auftretens-

²² Wie sich diskursanalytische Verfahren und *Culturomics* zueinander verhalten, bleibt offen. Ngram-Kurven bilden keine diskursiven Regelmässigkeiten ab. Michel Foucaults Diskurskonzept hebt allerdings auf Auftretenshäufigkeiten ab und ist mit Culturomics vermittelbar (vgl. dazu Sarasin 2012: 159–160).

²³ So der Untertitel von White 1991 [1978].

häufigkeiten bedarf gewiefter Informatiker, während sprachwissenschaftliche Kompetenz immer überflüssiger wird.

Gegen einen solchen alingustischen Reduktionismus schreiben gesellschaftshistorisch interessierte und kulturalistisch versierte KorpuslinguistInnen mit guten Gründen an (Linke 2003; vgl. auch Linke 2008). Eine Beschränkung auf „elektronische Zettelkästen“ (Bubenhofer 2018: 18) und *Datamining* in Texten, die als „unstrukturierte Daten“ (Bubenhofer 2018: 19) traktiert werden, führt mitnichten zu forschungsrelevanten Erkenntnissen, denn „eine theoretische Fundierung der Analysekategorien ist essentiell“ (Bubenhofer 2018: 24). Diese Einsicht verbindet heute kulturalistische Zugangsweisen, wie sie Angelika Linke vorschlägt, mit den erwähnten Versuchen, die Begriffsgeschichte zu aktualisieren. Hier gibt es mittlerweile nicht nur Projekte, die, nach wie vor durch Reinhart Koselleck inspiriert, Schlüsselbegriffe unter die Lupe nehmen, sondern auch Ansätze einer Mikrosemantik, die, vor allem für das Mittelalter und die Frühe Neuzeit, auf eine kleinteilige Analyse von sprachlich vermittelten Interaktionen in spezifischen Handlungssituationen abzielen (Kuchenbuch 2012; Schiel 2014; Schiel & Hanß 2014).

Zusammenfassend lässt sich festhalten, dass *Culturomics* und mit ihr die *Digital Humanities* insgesamt eine zwar technisch imponierende, jedoch inhaltlich armselige Veranstaltung bleiben, wenn sie nicht sprachphilosophisch, linguistisch und historisch fundiert bleiben. Das Staunen über große Textmassen und der technisch virtuose Umgang damit kann wissenschaftliche Neugier nicht ersetzen. Diese hängt nach wie vor von weiter gefassten Fragestellungen und theoretischen Vorentscheidungen ab. Besonders aussichtsreich scheint es, die „Great Digitization“ auf eine historische Anthropologie zu beziehen, welche von einer genuinen Historizität der menschlichen Natur ausgeht (vgl. Tanner 2004). Weiterführend sind auch Ansätze einer *Social Neuroscience*, die eine beträchtliche Plastizität des menschlichen Gehirns und eine Formbarkeit des Körpers unterstellen und danach fragen, wie sich die menschlichen Kulturtechniken in ein neuronales Substrat und einen korporalen Habitus einschreiben – statt umgekehrt zu vermuten, dass sie kulturelle Operatoren direkt aus der genetisch fixierten Architektur des Gehirns erklären lassen. In diesem Sinne schrieb der französische Historiker und Annales-Mitbegründer Lucien Febvre 1950, „die intellektuelle Kategorie, die wir in unseren Zerebralwerkstätten schmieden“, würde sich uns „mit derselben Wucht, derselben Tyrannei auf(drängen)“ wie die „Maschinen (...) aus Stahl“ und sie führe „ein weit zäheres Leben als die in unseren Fabriken hergestellte Mechanik“ (Febvre 1990: 211). Heute, nach dem „grossen Umzug der Welt in den Computer“ (Gugerli 2018: 198), müsste eine weniger schwerindustrielle Metaphorik gewählt werden, um die Wirklichkeit und damit die Wirkmächtigkeit der Sprache einsichtig zu machen.

5 Literatur

- Aebischer, Patrick (2018): Das Mittelalter lebt. Die digitale Revolution aus den Archiven von Venedig. Gastkolumne. *NZZ am Sonntag* 30. Juni 2018. <https://nzzas.nzz.ch/meinungen/das-mittelalter-lebt-die-digitale-revolution-aus-den-archiven-von-venedig-id.1399596> (23. August 2018).
- Bell, David A. (2018): Pity is treason. *The New York Review of Books* 28. Juni 2018, 75–76.
- Berry, Dave M. (2011): The computational turn. Thinking about the digital humanities. *Culture Machine* 12, 1–22.
- Bubenhof, Noah & Philipp Dreesen (2018): Linguistik als antifragile Disziplin? Optionen in der digitalen Transformation. *Digital Classics Online* 4 (1), 63–75.
- Bubenhof, Noah (2018): Wenn „Linguistik“ in „Korpuslinguistik“ bedeutungslos wird. Vier Thesen zur Zukunft der Korpuslinguistik. *Osnabrücker Beiträge zur Sprachtheorie* 92, 17–30.
- Bunz, Mercedes (2012): *Die stille Revolution. Wie Algorithmen Wissen, Arbeit, Öffentlichkeit und Politik verändern, ohne dabei viel Lärm zu machen*. Berlin: Suhrkamp.
- Chumtong, Jason & David Kaldewey (2017): Beyond the Google Ngram Viewer. Bibliographic databases and journal archives as tools for the quantitative analysis of scientific and meta-scientific concepts. *FIW Working Paper* 8. <https://www.fiw.uni-bonn.de/publikationen/FIWWorkingPaper/fiw-working-paper-no.-8> (23. August 2018).
- Crawford, Joseph (2013): *Gothic fiction and the invention of terrorism. The politics and aesthetics of fear in the age of the reign of terror*. London: Bloomsbury Academics.
- Diebolt, Claude & Michael Haupt (Hrsg.) (2016): *Handbook of cliometrics*. Berlin: Springer.
- Farge, Arlette (1989): *Le goût de l'archive*. Paris: Éditions du Seuil.
- Febvre, Lucien (1990): *Das Gewissen des Historikers*. Frankfurt am Main: Wagenbach.
- Geulen, Christian (2010): Plädoyer für eine Geschichte der Grundbegriffe des 20. Jahrhunderts. *Zeithistorische Forschungen* 7 (1), 79–97.
- Geulen, Christian (i. Dr.): *Zeitschleifen. Über Grundbegriffe des 20. Jahrhunderts*. München: Beck.
- Geulen, Christian & Christian Bermes (Hrsg.) (i. V.): *Politische Schlüsselbegriffe des 20. Jahrhunderts*.
- Gugerli, David (2018): *Wie die Welt in den Computer kam*. Frankfurt am Main: Fischer.
- Hagner, Michael (2015): *Zur Sache des Buches*. Göttingen: Wallstein.
- Harris, Randy Allen (1993): *The linguistics wars*. Oxford: Oxford University Press.
- Hölscher, Lucian (2002): *The theory and method of german ‚Begriffsgeschichte‘ and its impacts on the construction of an european political lexicon*. [http://www.ruhr-uni-bochum.de/lehrstuhl-ng3/publikationen/hoelscher/HistoryOfTheFuture\(2002\).pdf](http://www.ruhr-uni-bochum.de/lehrstuhl-ng3/publikationen/hoelscher/HistoryOfTheFuture(2002).pdf) (23. August 2018).
- Ilg, Yvonne & Theresa Schnedermann (2014): Schizophrenie und Burnout im Lichte sprachlicher Reflexion. In Ekkehard Felder & Marcus Müller (Hrsg.), *Diskurszukünfte. 10. Jahrestagung des Forschungsnetzwerks „Sprache und Wissen“*. *Jubiläumszeitschrift*, 37–39. Heidelberg: Universität Heidelberg. http://archiv.ub.uni-heidelberg.de/volltextserver/18444/1/Diskurszukunfte_Jubiläumszeitschrift.pdf (23. August 2018).
- Ilg, Yvonne & Anke Maatz (2015): Sprachliche Bilder von „Schizophrenie“ zwischen Fach- und Alltagssprache. In Daniel Sollberger et al. (Hrsg.): *Bilder der Schizophrenie*, 65–85. Berlin: Frank & Timme.
- Ilg, Yvonne (2015): Schizophrenie. Zur gemeinsprachlichen „Karriere“ eines Fachbegriffs und ihren Folgen. In Jianhua Zhu, Jin Zhao & Michael Szurawitzki (Hrsg.), *Germanistik zwischen Tradition und Innovation. Akten des XIII. Kongresses der Internationalen Vereinigung für Germanistik (IVG)*. Shanghai, 23.–30. 8. 2015. Bd. 2, 43–47. Frankfurt am Main: Lang.
- Ilg, Yvonne (i. V.): „Schizophrenie“ in der Alltagssprache. *Eine linguistische Begriffsgeschichte 1908–2009*.
- Jannidis, Fotis, Hubertus Kohle & Malte Rehbein (2017) (Hrsg.): *Digital Humanities. Eine Einführung*. Stuttgart: Metzler.
- Jensen, Kim Ebensgaard (2014): Linguistics and the digital humanities. (Computational) Corpus linguistics. *MedieKultur. Journal of media and communication research* 57, 115–134.
- Konersmann, Ralf (1994): *Der Schleier des Timanthes. Perspektiven der historischen Semantik*. Frankfurt am Main: Fischer.
- Koselleck, Reinhart (1972): Einleitung. In Otto Brunner, Werner Conze & Reinhart Koselleck (Hrsg.), *Geschichtliche Grundbegriffe. Historisches Lexikon zur politisch-sozialen Sprache in Deutschland*. Bd. 1, XXIII. Stuttgart: Klett-Cotta.
- Koselleck, Reinhart (1991): Einführung. In Hayden White: *Auch Klio dichtet oder Die Fiktion des Faktischen. Studie zur Tropologie des historischen Diskurses*, 1–6. Stuttgart: Klett-Cotta.
- Krämer, Sybille (2016): *Figuration, Anschauung, Erkenntnis. Grundlinien einer Diagrammatologie*. Berlin: Suhrkamp.
- Kuchenbuch, Ludolf (2012): *Reflexive Mediävistik. Textus – Opus – Feudalismus*. Frankfurt am Main: Campus.
- Leetaru, Kalev (2011): Culturomics 2.0. Forecasting large-scale human behavior using global news media tone. *First Monday* 16 (9). <http://firstmonday.org/ojs/index.php/fm/article/view/3663/3040> (23. August 2018).
- Linke, Angelika (2003): Sprachgeschichte – Gesellschaftsgeschichte – Kulturanalyse. In Helmut Henne, Horst Sitta & Herbert Ernst Wiegand (Hrsg.), *Germanistische Linguistik. Konturen eines Faches*, 25–65. Tübingen: Niemeyer.
- Linke, Angelika (2008): Kommunikation, Kultur und Vergesellschaftung. Überlegungen zu einer Kulturgeschichte der Kommunikation. In Ludwig M. Eichinger & Heidrun Kämper (Hrsg.), *Sprache – Kognition – Kultur. Sprache zwischen mentaler Struktur und kultureller Prägung*, 24–50. Berlin: De Gruyter.
- Lyons, John S. (2008): *Reflections on the cliometrics revolution. Conversations with economic historians*. London: Routledge.
- Manovich, Lev (2017): Cultural data. Possibilities and limitations of the digital data universe. In Oliver Grau (Hrsg.), *Museum and archive on the move. Changing cultural institutions in the digital era*, 259–276. Berlin: De Gruyter.
- McLuhan, Marshall (1962): *The Gutenberg galaxy. The making of typographic man*. London: Routledge.
- Michel, Jean-Baptiste et al. (2011): Quantitative analysis of culture using millions of digitized books. *Science* 331 (6014), 176–182.
- Müller, Ernst (2011): Verspätete Wirkung. Reinhart Kosellecks Begriffsgeschichte international. *Trajekte* 23, 22–25.
- Müller, Ernst & Falko Schmieder (2016): *Begriffsgeschichte und historische Semantik: ein kritisches Kompendium*. Berlin: Suhrkamp.
- Polastron, Lucien X. (2009): *The great digitization and the quest to know everything*. Rochester: Inner Traditions.

- Potthast, Martin et al. (2018): A plan for ancillary copyright. Original snippets. *CEUR-WS.org* 2079. <http://ceur-ws.org/Vol-2079/paper1.pdf> (23. August 2018).
- Reichert, Ramón (2014): *Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie*. Bielefeld: transcript.
- Rohmann, Gregor (2015): Worte – Konzepte – Bedeutungen. Welche Historische Semantik für das Mittelalter? Abschlussstagung des Leibniz-Projekts „Politische Sprache im Mittelalter. Semantische Zugänge“, 19.–20. Februar 2015, Frankfurt am Main [Tagungsbericht]. *H-Soz-Kult* 15. Mai 2015. <http://www.hsozkult.de/conferencereport/id/tagungsberichte-5972> (23. August 2018).
- Rosenberg, Scott (2017): How Google Book search got lost. *WIRED, backchannel* 4. November 2017. <https://www.wired.com/2017/04/how-google-book-search-got-lost/> (23. August 2018).
- Sarasin, Philipp (2012): Sozialgeschichte vs. Foucault im Google Books Ngram Viewer. Ein alter Streitfall in einem neuen Tool. In Pascal Maeder, Barbara Lüthi & Thomas Mergel (Hrsg.), *Wozu noch Sozialgeschichte? Eine Disziplin im Umbruch*, 151–174. Göttingen: Vandenhoeck & Ruprecht.
- Schechter, Ronald (2018): *A genealogy of terror in eighteenth-century France*. Chicago: University of Chicago Press.
- Schiel, Juliane & Stefan Hanß (2014): Semantics, practices and transcultural perspectives on mediterranean slavery. Semantiken, Praktiken und transkulturelle Perspektiven mediterraner Sklaverei. In Juliane Schiel & Stefan Hanß (Hrsg.), *Mediterranean Slavery Revisited*, 11–45. Zürich: Chronos.
- Schiel, Juliane (2014): Sklavennahme in der Seerepublik Venedig. Praktiken und Semantiken des Nehmens im spätmittelalterlichen Menschenhandel. *Geschichte in Wissenschaft und Unterricht* 65 (9/10), 586–599.
- Schönpflug, Wolfgang (1969): N-Gramm-Häufigkeiten in der deutschen Sprache. I. Monogramme und Digramme. *Zeitschrift für experimentelle und angewandte Psychologie* XVI, 157–183.
- Skinner, Quentin (2009): *Visionen des Politischen*. Berlin: Suhrkamp.
- Steinmetz, Willibald (2008): Vierzig Jahre Begriffsgeschichte. The Stare of the Art. In Heidrun Kämper & Ludwig M. Eichinger (Hrsg.), *Sprache – Kognition – Kultur. Sprache zwischen mentaler Struktur und kultureller Prägung*, 174–197. Berlin: De Gruyter.
- Tanner, Jakob (2004): *Historische Anthropologie zur Einführung*. Hamburg: Junius.
- Walther, Rudolf (1990): Terror/Terrorismus. In Otto Brunner, Werner Conze & Reinhart Koselleck (Hrsg.), *Geschichtliche Grundbegriffe. Historisches Lexikon zur politisch-sozialen Sprache in Deutschland*. Bd. 6, 323–444. Stuttgart: Klett-Cotta.
- White, Hayden (1991 [1978]): *Auch Klio dichtet oder Die Fiktion des Faktischen. Studie zur Topologie des historischen Diskurses*. Stuttgart: Klett-Cotta.

Reihe Germanistische Linguistik

Herausgegeben von
Mechthild Habermann und Heiko Hausendorf

Wissenschaftlicher Beirat
Karin Donhauser (Berlin), Stephan Elspaß (Salzburg),
Helmuth Feilke (Gießen), Jürg Fleischer (Marburg),
Stephan Habscheid (Siegen), Rüdiger Harnisch (Passau)

314

Linguistische Kulturanalyse

Herausgegeben von
Juliane Schröter, Susanne Tienken, Yvonne Ilg,
Joachim Scharloth und Noah Bubenhofer

DE GRUYTER